A-Collapsibility of Distribution Dependence and Quantile Regression Coefficients

Mark M. Meerschaert and P. Vellaisamy

Department of Statistics and Probability, Michigan State University, USA

and

Department of Mathematics, Indian Institute of Technology Bombay, India

Abstract

The Yule-Simpson paradox notes that an association between random variables can be reversed when averaged over a background variable. Cox and Wermuth introduced a new concept of distribution dependence between two random variables X and Y, and developed two dependence conditions, each of which guarantees that reversal cannot occur. Ma, Xie and Geng studied the collapsibility of distribution dependence over a backround variable W, under a rather strong homogeneity condition. Collapsibility ensures the association remains the same for conditional and marginal models, so that Yule-Simpson reversal cannot occur. In this paper, we investigate a more general notion called A-collapsibility. The conditions of Cox and Wermuth imply A-collapsibility, without assuming homogeneity. In fact, we show that, when W is a binary variable, collapsibility is equivalent to A-collapsibility plus homogeneity, and A-collapsibility is equivalent to the conditions of Cox and Wermuth. Recently, Cox extended Cochran's result on regression coefficients of conditional and marginal models, to quantile regression coefficients. The conditions of Cox and Wermuth are also sufficient for A-collapsibility of quantile regression coefficients. Under a conditional completeness assumption, they are also necessary.

1 Introduction

There are several ways to interpret the association between a response and an explanatory variable. The measure of association may be measured by odds ratio or relative risk or interaction parameters of the corresponding log-linear model (for categorical variables), regression coefficient or distribution dependence (for continuous variables). The concept of collapsibility with respect to these parameters was well studied by Bishop (1971), Cox (2003), Cox and Wermuth (2003), Geng (1992), Ma *et. al.* (2006), Vellaisamy and Vijay (2007, 2008), Wermuth (1987,

1989) and Whittemore (1978), among others. Cox and Wermuth (2003) defined distribution dependence as a measure of association between two variables, and discussed the reversal effect when a background variable (sometimes unobserved) is condensed. They obtained sufficient conditions for no reversal effect, that is, for the non-occurence of Yule and Simpson's paradox. Recently, Ma *et. al.* (2006) proved that the conditions of Cox and Wermuth (2003) are indeed necessary and sufficient for collapsibility of distribution dependence, under the assumption that distribution dependence is homogeneous over the background variable. Homogeneity is a rather strong assumption, and restricts the applicability of these conditions.

The concept of A-collapsibility for random coefficient models was introduced and discussed in Vellaisamy and Vijay (2008). In the same spirit, this paper considers (average)A-collapsibility of distribution dependence. A-collapsibility means simply that the conditional effect averages over the background variable to the corresponding marginal effect. The conditions of Cox and Wermuth (2003) are shown to be sufficient for A-collapsibility, and also necessary when W is a binary variable. A necessary condition for A-collapsibility in terms of conditional densities is also obtained. Recently, Cox extended Cochran's result on regression coefficients of conditional and marginal models, to quantile regression coefficients. The conditions of Cox and Wermuth are also shown to be sufficient for A-collapsibility of quantile regression coefficients. Under a conditional completeness assumption, they are even necessary.

2   Collapsibility of Distribution Dependence

Let X and Y be two random variables. The dependence Y on X is called stochastically increasing if $P(Y > y \mid X = x)$ is increasing in x for all y. When Y is continuous, this is equivalent to saying that the conditional distribution function $G(y \mid x)$ decreasing in x, that is,

$$\frac{\partial G(y \mid x)}{\partial x} \leq 0, \tag{2.1}$$

for all y and x, with strict inequality in a region of positive probability. When X is discrete, the partial differentiation is replaced by differencing between adjacent levels of X. Assume now, for simplicity, that the variables $Y, X,$ and W are continuous. Suppose also that Y given $X = x$ and $W = w$ is stochastically increasing in x for all w, so that $\dfrac{\partial F(y \mid x, w)}{\partial x} \leq 0$ for all y, x and w. Then,

$$F(y \mid x) = P(Y \leq y \mid X = x) = \int F(y \mid x, w) f(w \mid x) \, dw.$$

On differentiating with respect to x, we have

$$\frac{\partial F(y \mid x)}{\partial x} = \int \frac{\partial F(y \mid x, w)}{\partial x} f(w \mid x) \, dw + \int F(y \mid x, w) \frac{\partial f(w \mid x)}{\partial x} \, dw. \tag{2.2}$$

If $X \perp W$, then $f(w \mid x) = f(w)$ and so (Cox (2003))

$$\frac{\partial f(w \mid x)}{\partial x} = 0,$$

leading to,

$$\frac{\partial F(y \mid x)}{\partial x} = \int \frac{\partial F(y \mid x, w)}{\partial x} f(w) \, dw. \tag{2.3}$$

Thus, when $X \perp W$, we have from (2.3),

$$\frac{\partial F(y \mid x, w)}{\partial x} \leq 0 \implies \frac{\partial F(y \mid x)}{\partial x} \leq 0, \text{ for all } y, x \text{ and } w.$$

Thus, Y remains stochastically increasing in x after marginalization over the covariate W. Note in general (see (2.2)) it is possible that $\frac{\partial F(y \mid x, w)}{\partial x} \leq 0$, for all y, x and w but $\frac{\partial F(y \mid x)}{\partial x} > 0$ for some y and x, implying the reversal effect. That is, the dependence of Y and X is no longer stochastically increasing. This reversal effect is known as Yule-Simpson paradox (e.g., see Cox and Wermuth (2003)).

Let Y be a response variable, X be an explanatory variable and W be a background variable. The function $\frac{\partial F(y \mid x, w)}{\partial x}$ is called a distribution dependence function. If the variable X is categorical with support $S(X) = \{1, \cdots, I\}$, then the distribution dependence function is defined as (see Cox (2003)).

$$\frac{\partial F(y \mid x, w)}{\partial x} = \Delta_x F(y \mid i, w) = P(Y \leq y \mid i + 1, w) - P(Y \leq y \mid i, w), \tag{2.4}$$

for $i = 1, 2, \cdots, I - 1$.

The following definitions are due to Ma *et. al.* (2006).

Definition 2.1 *The distribution dependence function is said to be homogeneous with respect to W if*

$$\frac{\partial F(y \mid x, w)}{\partial x} = \frac{\partial F(y \mid x, w')}{\partial x},$$

for all y, x and $w \neq w'$.

Definition 2.2 *The distribution dependence function is said to be collapsible over W if*

$$\frac{\partial F(y \mid x, w)}{\partial x} = \frac{\partial F(y \mid x)}{\partial x}, \text{ for all } y, x \text{ and } w,$$

and uniformly collapsible if

$$\frac{\partial F(y \mid x, W \in A)}{\partial x} = \frac{\partial F(y \mid x)}{\partial x}$$

for all y, x and A in the support of W. When W is ordinal, the set A is of the form $(i, i + 1, \cdots, i + j)$.

Note that uniformly collapsible implies collapsible, and collapsible implies homogeneous. Homogeneity is commonly assumed for pooled estimation as in Mantel and Haenszel (1959). Ma *et. al.* (2006) showed that the distribution dependence function is uniformly collapsible iff either: (a) $Y \perp X|W$; or (b) $X \perp W$ and $\dfrac{\partial F(y|x,w)}{\partial x}$ is homogeneous in w. Cox and Wermuth (2003) note that either condition (a) or (b) is sufficient to ensure that no reversal effect can occur when marginalizing the background variable W.

3 A-Collapsibility of Distribution Dependence

A-collapsibility is a weaker criterion for non-reversal. It requires only that the conditional effect averages over the background variable to the corresponding marginal effect. In particular, it does not assume homogeneity. In fact, collapsibility is equivalent to A-collapsibility plus homogeneity (see Theorem 3.3), when W is a binary variable. Homogeneity is a strong assumption (see Ma et. al. (2006), p.129). Most of the models that are encountered in practice are not homogeneous (e.g., see example 3.1). As an another example, consider a simple non-linear regression

$$Y = m(X, W) + \epsilon,$$

where $m(x, w) = \alpha_1 x + \alpha_2 w + \alpha_3 x w$, and $\epsilon \sim N(0, \sigma^2)$. Then,

$$\frac{\partial F(y|x,w)}{\partial x} = (\alpha_1 + \alpha_3 w) \phi \left(\frac{y - m(x,w)}{\sigma} \right),$$

where ϕ is the standard normal density, so that this example is not homogeneous over W.

The above observations motivate our definition of (average) A-collapsibility of distribution dependence, similar to A-collapsibility of regression coefficients described in Vellaisamy and Vijay (2008). Indeed, this seems to be a natural definition of collapsibility for a large class of conditional distribution functions.

Definition 3.1 *The distribution dependence function* $\dfrac{\partial F(y|x,w)}{\partial x}$ *is (average)A-collapsible over W if*

$$E_{W|X=x} \left(\frac{\partial F(y|x,W)}{\partial x} \right) = \frac{\partial F(y|x)}{\partial x}, \text{ for all } y \text{ and } x. \tag{3.1}$$

Note that the above definition is a natural extension of simple collapsibility of distribution dependence. Indeed, when $\dfrac{\partial F(y|x,w)}{\partial x}$ is homogeneous over W, A-collapsibility reduces to collapsibility. The next result shows that the conditions of Cox and Wermuth (2003) are sufficient for A-collapsibility.

Theorem 3.1 *(a): Either of the conditions*

(i) $Y \perp W \mid X$; or

(ii) $W \perp X$

are sufficient for the distribution dependence function $\dfrac{\partial F(y|x,w)}{\partial x}$ to be A-collapsible over a discrete background variable W.

(b): Conversely, if W is binary, say $w \in \{1,2\}$, then the condition (i) or (ii) is also necessary.

Next we provide an example that is A-collapsible, but neither collapsible nor homogeneous.

Example 3.1 *Consider the following $2 \times 2 \times 2$ table.*

		W	
Y	X	1	2
1	1	25	35
	2	75	60
2	1	35	15
	2	45	40

Here, we have

$$\Delta_x F(1|1,1) = P(Y = 1|X = 2, W = 1) - P(Y = 1|X = 1, W = 1) = 0.208; \quad \text{and}$$

$$\Delta_x F(1|1,2) = P(Y = 1|X = 2, W = 2) - P(Y = 1|X = 1, W = 2) = -0.1.$$

That is, the distribution dependence is not homogeneous. Also, note from the marginal table of Y and X,

$$\Delta_x F(1|1) = P(Y = 1|X = 2) - P(Y = 1|X = 1) = 0.068 \neq \Delta_x F(1|1, w),$$

so that the distribution dependence function is not collapsible over W. However, from the marginal table of X and W,

		W	
		1	2
X	1	60	50
	2	120	100

it can be seen that $X \perp W$ and

$$
\begin{aligned}
E_{W|X=1}\left(\Delta_x F(1|1, W)\right) &= \sum_w (\Delta_x F(1|1, w)) f_{W|X}(w|x) \\
&= \Delta_x F(1|1, 1) f_W(1) + \Delta_x F(1|1, 2) f_W(2) = .068 \\
&= \Delta_x F(1|1).
\end{aligned}
$$

Therefore, the distribution dependence function is A-collapsible with respect to the background variable W.

Remark 3.1 *Theorem 1 in Ma et. al. (2006) implies that $Y \perp W \mid X$ is a sufficient condition for collapsibility. However, one can easily see that if $Y \perp W \mid X$, then*

$$F(y|x,w) = F(y|x) = F(y|x,w'), \forall\ w \neq w'. \tag{3.2}$$

Hence,

$$\frac{\partial F(y|x,w)}{\partial x} = \frac{\partial F(y|x)}{\partial x} = \frac{\partial F(y|x,w')}{\partial x} \tag{3.3}$$

implying that the homogeneity condition is also satisfied. Thus, $Y \perp W \mid X \implies$ homogeneity of distribution dependence. So, Theorem 1 in Ma et. al. (2006) pertains only to the class of distribution dependence functions that are homogeneous over W.

Note also from (2.2) that A-collapsibility holds if and only if

$$\int F(y \mid x, w) \frac{\partial f(w \mid x)}{\partial x}\, dw = 0 \ \ \text{for all } (y, x). \tag{3.4}$$

The following example shows that A-collapsibility can hold even when neither condition (i) nor condition (ii) of Theorem 3.1 hold. Hence these conditions are not necessary, unless the background variable W is binary.

Example 3.2 Let Y, given $X = x$ and $W = w$, follow uniform $U(0, (x^2 + (w - x)^2)^{-1})$ so that

$$F(y|x,w) = y(x^2 + (w - x)^2), \ \ 0 < y < (x^2 + (w - x)^2)^{-1}. \tag{3.5}$$

Assume also $(W|X = x) \sim N(x, 1)$ so that

$$\frac{\partial}{\partial x} f(w|x) = -\phi'(w - x) = (w - x)\phi(w - x), \tag{3.6}$$

where $\phi(z)$ denotes the density of $N(0, 1)$ distribution. Hence,

$$
\begin{aligned}
\int F(y|x,w) \frac{\partial}{\partial x} f(w|x) dx &= y \int (x^2 + (w - x)^2)(w - x)\phi(w - x) dw \\
&= y[x^2 \int (w - x)\phi(w - x) dw + \int (w - x)^3 \phi(w - x) dw] \\
&= y[x^2 \int t\phi(t) dt + \int t^3 \phi(t) dt] \\
&= 0, \ \ \text{for all } (y, x) \in S_{yx}. \tag{3.7}
\end{aligned}
$$

Thus, from (3.4), A-collapsibility over W holds, but neither condition (i) nor condition (ii) is satisfied.

The following result provides a necessary condition for A-collapsibility. It also shows that A- collapsibility of distribution dependence implies A-collapsibility of density dependence.

Theorem 3.2 *Suppose $F(y|x,w)$ and $F(y|x)$ admit continuous mixed partial derivatives (with respect to y and x). Then a necessary condition for A-collapsibility of the distribution dependence function over W is*

$$E_{W|X=x}\left(\frac{\partial f(y|x,W)}{\partial x}\right) = \frac{\partial f(y|x)}{\partial x}, \quad \forall \ (y,x). \tag{3.8}$$

Remark 3.2 *(i)* Note in Example 3.2,

$$\frac{\partial f(y|x,w)}{\partial x} = 2x + 2(x-w); \quad f(w|x) = \phi(w-x).$$

Hence,

$$E_{W|x}(\frac{\partial f(y|x,W)}{\partial x}) = 2x + 2(x - E(W|x)) = 2x.$$

Also,

$$f(y|x) = \int f(y|x,w)f(w|x)dw \tag{3.9}$$

$$= \int (x^2 + (w-x)^2)\phi(w-x)dw \tag{3.10}$$

$$= x^2 + 1. \tag{3.11}$$

That is, $(Y|x) \sim U(0,(1+x^2)^{-1})$. Thus, the necessary condition (3.8) is verified.

The following result shows that collapsibility implies A-collapsibility plus homogeneity. If the background variable W is binary, then collapsibility equals A-collapsibility plus homogeneity.

Theorem 3.3 *Let (i) C_H^W, (ii) C^W and (iii) C_A^W respectively denote the class of distribution functions with distribution dependence (i) homogeneous over W, (ii) collapsible over W and (iii) A-collapsible over W. Then*

$$C^W \subseteq C_H^W \cap C_A^W,$$

and the equality holds when W is binary.

4   A-Collapsibility of Quantile Regression Coefficients

Assume, for brevity, the random variables Y, X and W are continuous with finite variances. Cochran (1938) proved the following relation for linear regression coefficients:

$$\beta_{yx} = \beta_{yx.w} + \beta_{yw.x}\beta_{wx}, \tag{4.1}$$

where β_{yx} denotes the linear regression coefficient of Y on X, and $\beta_{yx.w}$ denotes corresponding coefficient of Y on X and W, and so forth. Equation (4.1) decomposes the effect of a unit change in X on the response variable Y into two parts, the first being the effect with W fixed, and the second a product of two effects: The effect of a unit change in X on the moderating variable W, times the effect of a unit change in W on the response Y when X is fixed. Cox (2007) notes that (4.1) is essentially the formula for the total derivative

$$\frac{dy}{dx} = \frac{\partial y}{\partial x} + \frac{\partial y}{\partial w}\frac{dw}{dx}$$

and hence could be extended to the more general setting of quantile regression coefficients, which we now describe. Given $0 < \eta < 1$, the function $y_\eta = y_\eta(x)$ satisfying $F(y_\eta|x) = \eta$ is called η-th quantile function. The function

$$q_x(y|x) = \frac{-\frac{\partial}{\partial x}F(y|x)}{f(y|x)} \tag{4.2}$$

is called the quantile regression coefficient (see equation (2) of Cox (2007)). Note that

$$\frac{\partial}{\partial x}y_\eta(x) = q_x(y_\eta(x)|x)$$

by implicit differentiation. Hence, the quantile regression function describes the effect of a unit change in X on quantiles of Y. Similarly,

$$q_x(y|x,w) = \frac{-\frac{\partial}{\partial x}F(y|x,w)}{f(y|x,w)} \tag{4.3}$$

represents the conditional quantile regression coefficient. Cox (2007) p.757 established that

$$q_x(y|x) = E_{W|y,x}\{\delta(y|x,W)\}, \tag{4.4}$$

where $\delta(y|x,w) = q_x(y|x,w) + q_w(y|x,w)q_x(w|x)$ represents the total effect on quantiles of Y of a unit change in X, calculated at (x,w). When $\delta(y|x,w)$ does not depend on w, Cox (2007) noted that

$$q_x(y|x) = \delta(y|x,w), \tag{4.5}$$

a result similar to that of Cochran (1938). Our interest lies in the quantile regression coefficients $q_x(y|x)$ and $q_x(y|x,w)$.

Definition 4.1 *The quantile regression coefficient $q_x(y|x,w)$ is A-collapsible over W if*

$$q_x(y|x) = E_{W|y,x}(q_x(y|x,W)). \tag{4.6}$$

The next result shows that the conditions (i) and (ii) of Cox and Wermuth (2003) are sufficient for A-collapsibility.

Theorem 4.1 *The quantile regression coefficient $q_x(y|x, w)$ is A-collapsible over W if (i) $Y \perp W|X$ or (ii) $W \perp X$.*

We next show, in general, that the converse of Theorem 4.1 is not true. Let S_{yx} denote the support of (Y, X). Note from (A.13), A-collapsibility holds

$$\iff \int q_w(y|x, w)q_x(w|x)dF(w|y, x) = 0, \ \forall \ (y, x) \in S_{yx} \tag{4.7}$$

$$\iff \int (q_w(y|x, w)q_x(w|x))f(y|x, w)\frac{f(w|x)}{f(y|x)}dw = 0$$

$$\iff \int F_w(y|x, w)F_x(w|x)dw = 0, \ \forall \ (y, x) \in S_{yx}. \tag{4.8}$$

Example 4.1 Let $X > 0$ and W be real-valued continuous random variables with

$$F(w|x) = \Phi(\frac{w}{x}), \ x > 0, \ w \in \mathbb{R},$$

so that

$$F_x(w|x) = -\frac{w}{x^2}\phi(\frac{w}{x}),$$

where Φ and ϕ denote respectively the distribution and the density function of $Z \sim N(0, 1)$. Also, let

$$F(y|x, w) = \frac{y + x - w}{2x}, \ w - x < y < w + x,$$

so that Y, given $X = x$ and $W = w$, follows uniform $U(w - x, \ w + x)$, and

$$F_w(y|x, w) = -\frac{1}{2x}, \ w - x < y < w + x.$$

Also,

$$\int_{-\infty}^{\infty} F_w(y|x, w)F_x(w|x)dw = \frac{1}{2x^2} \int_{-\infty}^{\infty} \frac{w}{x}\phi(\frac{w}{x})dw$$

$$= \frac{1}{2x} \int_{-\infty}^{\infty} t\phi(t)dt$$

$$= 0, \ \text{for all } (y, x) \in S_{yx}.$$

Then from (4.8) A-collapsibility holds, but neither condition (i) nor condition (ii) is satisfied.

Finally, we identify a class of distributions for which condition (i) or condition (ii) is also necessary.

Definition 4.2 *The random variable W is said to be conditionally complete if $\{F(w|y, x)|(y, x) \in S_{yx}\}$ is complete, that is,*

$$E_{W|y,x}[h(W)] = 0 \Longrightarrow h(W) = 0 \ a.e. \ \mathcal{P}_{W|y,x}.$$

Theorem 4.2 *Let $\{F(w|y,x)|(y,x) \in S_{yx}\}$ be complete. Then condition (i) or (ii) of Theorem 4.1 is also necessary.*

Observe that if the densities $f(w|y,x)$ belong to exponential family, then $\{F(w|y,x)\}$ is complete.

Appendix A: Proofs

Proof of Theorem 3.1. Let Y and X be continuous and W be discrete. First assume Condition (i) holds. Then

$$E_{W|X=x}\left(\frac{\partial F(y|x,W)}{\partial x}\right) = E_{W|X=x}\left(\frac{\partial F(y|x)}{\partial x}\right) = \frac{\partial F(y|x)}{\partial x}$$

and hence A-collapsibility holds. Assume next Condition (ii) holds. Then

$$
\begin{aligned}
\frac{\partial F(y|x)}{\partial x} &= \frac{\partial}{\partial x}\left[\sum_w F(y|x,w)f_{W|X}(w|x)\right] \\
&= \sum_w \left(\frac{\partial}{\partial x}F(y|x,w)\right)f_W(w) \\
&= \sum_w \left(\frac{\partial F(y|x,w)}{\partial x}\right)f_{W|X}(w|x) \\
&= E_{W|X=x}\left(\frac{\partial F(y|x,W)}{\partial x}\right),
\end{aligned}
$$

showing again that A-collapsibility holds. The proof for W continuous is similar.

As to the converse, let

$$E_{W|X=x}\left(\frac{\partial F(y|x,W)}{\partial x}\right) = \frac{\partial F(y|x)}{\partial x}$$

hold for all y and x. Then,

$$
\begin{aligned}
\sum_w \left(\frac{\partial F(y|x,w)}{\partial x}\right)f_{W|X}(w|x) &= \frac{\partial}{\partial x}\left\{\sum_w F(y|x,w)f_{W|X}(w|x)\right\} \\
&= \sum_w f_{W|X}(w|x)\frac{\partial}{\partial x}F(y|x,w) \\
&\quad + \sum_w F(y|x,w)\frac{\partial}{\partial x}f_{W|X}(w|x). \qquad \text{(A.1)}
\end{aligned}
$$

Hence,

$$\sum_w F(y|x,w)\frac{\partial}{\partial x}f_{W|X}(w|x) = 0, \text{ for all } x,y. \qquad \text{(A.2)}$$

Since $w \in \{1, 2\}$ is binary, we have

$$\frac{\partial}{\partial x} f_{W|X}(1|x) = f_{W|X}(1|x) - f_{W|X}(2|x)$$

$$\frac{\partial}{\partial x} f_{W|X}(2|x) = f_{W|X}(2|x) - f_{W|X}(1|x)$$

and hence we get from (A.2),

$$\{F(y|x, 1) - F(y|x, 2)\}\frac{\partial}{\partial x} f_{W|X}(1|x) = 0, \text{ for all } y \text{ and } x.$$

Thus, we get $F(y|x, 1) = F(y|x, 2)$ or $\frac{\partial}{\partial x} f_{W|X}(1|x) = 0$, which are equivalent, respectively, to

$$Y \perp W \mid X \quad \text{or} \quad X \perp W.$$

To see that $\frac{\partial}{\partial x} f_{W|X}(1|x) = 0$ implies $X \perp W$, note that if $P(W = 2|X = x) = P(W = 1|X = x)$ for all x, then $P(W = 2, X = x) = P(W = 1, X = x) = 0.5P(X = x)$ for all x, and $X \perp W$ follows easily.

Proof of Theorem 3.2. We give the proof for the case of discrete W. Assume A-collapsibility holds. Then from (A.2),

$$\sum_w F(y|x, w)\frac{\partial}{\partial x} f_{W|X}(w|x) = 0, \text{ for all } x, y. \tag{A.3}$$

Also,

$$\sum_w F(y|x, w) f(w|x) = F(y|x), \ \forall \ (y, x). \tag{A.4}$$

Differentiating (A.4) with respect to x and using (A.3), we get

$$\sum_w \frac{\partial F(y|x, w)}{\partial x} f(w|x) = \frac{\partial F(y|x)}{\partial x}, \ \forall \ (y, x). \tag{A.5}$$

Differentiating (A.5) now with respect to y, we get

$$\sum_w \frac{\partial^2 F(y|x, w)}{\partial y \partial x} f(w|x) = \frac{\partial^2 F(y|x)}{\partial y \partial x}, \ \forall \ (y, x). \tag{A.6}$$

Since $F(y|x)$ has continuous mixed partial derivatives, we have

$$\frac{\partial^2}{\partial y \partial x} F(y|x) = \frac{\partial^2}{\partial x \partial y} F(y|x) = \frac{\partial}{\partial x} f(y|x)$$

(e.g., see Apostal (1962) p. 214), and hence

$$\frac{\partial^2 F(y|x)}{\partial y \partial x} = \frac{\partial f(y|x)}{\partial x}; \quad \frac{\partial^2 F(y|x, w)}{\partial y \partial x} = \frac{\partial f(y|x, w)}{\partial x} \ \forall \ (y, x).$$

11

Substituting the above facts in (A.6),

$$\sum_w \frac{\partial f(y|x,w)}{\partial x} f(w|x) = \frac{\partial f(y|x)}{\partial x} \ \forall \ (y,x), \tag{A.7}$$

which proves the result.

Proof of Theorem 3.3. Let C_1 and C_2 denote respectively the class of distribution functions that satisfy the conditions (i) and (ii) of Theorem 3.1. Then from Theorem 3.1,

$$C_1 \cup C_2 \subseteq C_A^W. \tag{A.8}$$

Suppose now $F \in C^W$ so that

$$\frac{\partial F(y|x,w)}{\partial x} = \frac{\partial F(y|x)}{\partial x}, \ \forall \ (y,x,w).$$

This implies

$$
\begin{aligned}
E_{W|X=x}\left(\frac{\partial F(y|x,W)}{\partial x}\right) &= E_{W|X=x}\left(\frac{\partial F(y|x)}{\partial x}\right) \\
&= \frac{\partial F(y|x)}{\partial x},
\end{aligned}
$$

and hence A-collapsibility holds. Thus, $C^W \subset C_A^W$. Also, by the definition of collapsibility, $C^W \subset C_H^W$ and so

$$C^W \subseteq C_H^W \cap C_A^W. \tag{A.9}$$

By Remark 3.1 and Theorem 1 of Ma et al. (2006), if $F \in C_H^W \cap [C_1 \cup C_2]$, then $F \in C^W$, so that

$$C_H^W \cap C_1 \cup C_2 \subseteq C^W. \tag{A.10}$$

Let now W be binary. Then by Theorem 3.1, $C_1 \cup C_2 = C_A$, and so

$$C_A^W \cap C_A \subseteq C^W. \tag{A.11}$$

The result now follows from (A.9) and (A.11).

Proof of Theorem 4.1. From Cox's result (4.4),

$$
\begin{aligned}
q_x(y|x) &= E_{W|y,x}(q_x(y|x,W)) \tag{A.12} \\
&\iff E_{W|y,x}(q_w(y|x,W)q_x(W|x)) = 0 \\
&\iff \int (q_w(y|x,w)q_x(w|x))dF(w|y,x) = 0, \text{ for all } (y,x). \tag{A.13}
\end{aligned}
$$

If Condition (i) holds, then since

$$Y \perp W|X \iff F(y|x,w) = F(y|x) \text{ for all } y, \ x \text{ and } w, \tag{A.14}$$

we have $q_w(y|x, w) = 0$. Hence, (A.12) holds.

If Condition (ii) $W \perp X$ holds, then,

$$
\begin{aligned}
F(w|x) &= F(w) \text{ for all } (w, x) \\
&\Rightarrow q_x(w|x) = 0 \text{ for all } (w, x),
\end{aligned}
$$

which in turn proves (A.12). This proves the result.

Proof of Theorem 4.2. Let A-collapsibility hold. Then from (4.7),

$$
\int q_w(y|x, w) q_x(w|x) dF(w|y, x) = 0, \text{ for all } (y, x) \in S_{yx}.
$$

The conditional completeness now implies

$$
q_w(y|x, w) q_x(w|x) = 0, \text{ for all } (y, x) \in S_{yx} \tag{A.15}
$$

which is equivalent to

$$
q_w(y|x, w) = 0, \text{ or } q_x(w|x) = 0.
$$

That is, condition (i) or (ii) holds.

References

Apostol, T. M. (1962). *Calculus.* Vol. II. Blaisdell Publishing Company, New York.

Bishop, Y. M. M. (1971). Effects of collapsing multidimensional contingency tables. *Biometrics*, **27**, 545-562.

Cochran, W. G. (1938). The omission or addition of an independent varaible in multiple linear regression. *J.R. Statist. Soc. Suppl.*, **5**, 171-176.

Cox, D. R. (2007). On a generalization of a result of W.G. Cochran. *Biometrika*, **94**, 755-759.

Cox, D. R. (2003). Conditional and marginal association for binary random variables. *Biometrika*, **90**, 982-984.

Cox, D. R. and Wermuth, N. (2003). A general condition for avoiding effect reversal after marginalization. *J. R. Statist. Soc. B*, **65**, 937-941.

Geng, Z. (1992). Collapsibility of relative risk in contingency tables with a response variable. *J. R. Statist. Soc. B*, **54**, 585-593.

Koenker, R. (2005). Quantile Regression. Cambridge University Press, Cambridge.

Ma, Z., Xie, X. and Geng, Z. (2006). Collapsibility of distribution dependence. *J. R. Statist. Soc. B* , **68**, 127-133.

Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natn. Cancer Inst.*, **22**, 719748.

Vellaisamy, P. and Vijay, V. (2007). Some collapsibility results for n-dimensional contingency tables. *Ann. Inst. Statist. Math.*, **59**, 557-576.

Vellaisamy, P. and Vijay, V. (2008). Collapsibility of regression coefficients and its extensions. *J. Statist. Plann. Inference*, **138**, 982-994.

Wermuth, N. (1989). Moderating effects of subgroups in linear models. *Biometrika*, **76**, 81-92.

Wermuth, N. (1987). Parametric collapsibility and the lack of moderating effects in contingency tables with a dichotomous response variable. *J. R. Statist. Soc. B*, **49**, 353-364.

Whittemore, A. S. (1978). Collapsibility of multidimensional contingency tables. *J. R. Statist. Soc. B*, **40**, 328-340.

August 23, 2010